

Section 2. Creating and Using GIS data#

2.1 Data types: Vector and Raster models#

Spatial data can be most simply defined as information that describes the distribution of things upon the surface of the earth. In effect any information concerning the location, shape of, and relationships among, geographic features (Walker 1993; DeMers 1997). In archaeology we routinely deal with an enormous amount of spatial data, varying in scale from the relative locations of archaeological sites upon a continental landmass down to the positions of individual artefacts within an excavated context. The first half of this section highlights the most important issues that need to be considered in incorporating common sources of spatial data within a GIS database. It comprises a short review of the particular issues that relate to obtaining and integrating spatial data within the GIS database. This concentrates on generic concerns such as projections, precision, accuracy and scale and is followed by a consideration of more source-specific issues. Throughout, the emphasis is upon the importance of carefully recording information about the various data themes.

There are two principal GIS data-models in widespread use, which are termed vector and raster. They differ in how they conceptualise, store and represent the spatial locations of objects.

It should be noted that, to date, the principal applications of GIS within archaeology have been restricted to 2-dimensional models, and at best 2.5 dimensional representations. The latter are a result of the inability of currently available analysis and display tools to adequately deal with truly 3-dimensional data. As a direct result, the issues we will be discussing here are concerned solely with the integration, management, analysis and archiving of representations of 2/2.5-dimensional space.

2.1.1 The Vector Model

In the vector model, the spatial locations of features are defined on the basis of co-ordinate pairs. These can be discrete, taking the form of points (POINT or NODE data); linked together to form discrete sections of line (ARC or LINE data); linked together to form closed boundaries encompassing an area (AREA or POLYGON data). Attribute data pertaining to the individual spatial features is maintained in an external database.

Figure 1: Vector and Raster Data. Cottam Project geophysics data displayed as a raster background and overlaid with an aerial photograph interpretation (blue line data), coins (red point data), and metal artefacts (black point data). Figure created by Peter Halls using data from the Cottam Project directed by Julian Richards. Image copyright © Archaeology Data Service.

In dealing with vector data an important concept is that of topology. Topology, derived from geometrical mathematics, is concerned with order, contiguity and relative position rather than with actual linear dimensions. A good illustration of a topological map is that of the London Underground metro system. This well-known map is a precise representation of the stations (points or nodes) and the routes (arcs or lines) between them, yet provides only a very approximate indication of their relative locations and no indication of distances between them.

Topology is useful in GIS because many spatial modelling operations do not require co-ordinate locations, only topological information -- for example to find an optimal path between two points requires a list of the arcs or lines that connect to each other and the cost to traverse them in each direction. It is also possible to perform the same spatial modelling and interrogation processes without using stored topology, by processing the geometrical data directly, as in such GIS as ArcMap and MapInfo, or by generating topology *on the fly*, as and when it is required. The latter is the approach taken by Intergraph, amongst other major GIS suppliers.

For a detailed discussion of the vector model see Aronoff 1989 and Burrough 1986.

Important information to record about vector files

The following information should always be recorded when assembling, compiling and utilising vector data:

- The data type: Point, Line or Polygon
- Type of topology which the file contains, such as line, network, closed area or arc-node
- Details of any automatic vector processing applied to the theme (such as snap-to-nearest-node)
- State of the topology in the file, particularly whether it is 'clean' (topologically consistent) or contains inconsistencies that may require further intervention or processing. This is particularly important where arc-node data is concerned
- Projection system
- Co-ordinate system

2.1.2 The Raster Model

Here the spatial representation of an object and its related non-spatial attribute are merged into a unified data file. In practice the area under study is covered by a fine mesh, or matrix, of grid cells and the particular ground surface attribute value of interest occurring at the centre of each cell point is recorded as the value for that cell. It should be noted that whilst some raster models support the assignment of values to multiple attributes per discrete cell, others adhere strictly to a single attribute per cell structure.

Within this model spatial data is not continuous but is divided into discrete units. In terms of recording where individual cells are located in space, each is referenced according to its row and column position within the overall grid. To fix the relative spatial position of the overall grid, i.e. to geo-reference it, the four corners are assigned planar co-ordinates. An important concept concerns the size of the component grid cells and is referred to as grid-resolution. The finer the resolution the more detailed and potentially closer to ground truth a raster representation becomes.

Unlike the vector model there are no implicit topological relationships in the data, we are after all not recording individual spatial features but instead the behaviour of attributes in space. For a detailed discussion of the raster model see Aronoff 1989 and Burrough 1986.

Important information to record about raster files

The following information should always be recorded when assembling, compiling and utilising raster data:

- grid size (number of rows and columns)
- grid resolution
- georeferencing information, e.g. corner co-ordinates, source projection.

2.1.3 Choice of Vector, Raster or Combined Forms of Spatial Database

The choice of vector, raster, or combined, forms for the spatial database may be determined by the GIS software in use and its ability to manipulate certain types of data.

Vector means of managing and manipulating the data are to be preferred for handling information relating to discrete points, delimited boundaries, alignment of linear features, etc. Thus a vector model would be used for storing, and manipulating, an excavation plan.

Raster means of managing and manipulating the data are to be preferred for handling continuous information such as altitude (see Digital Elevation Models, below), vegetation, etc., and are the digital

form in which information from Geophysical Survey, Aerial Photography, and other forms of Remote Sensing and non-invasive survey, are delivered.

Where both data types are required to be used *together* a GIS capable of manipulating both is required. When combining and integrating information from a variety of sources the following points should be kept in mind:

- All spatial data must be recorded in the same co-ordinate system. Data which are recorded to some other system must be transformed/projected to the required co-ordinate system.
- All spatial data should be to the same *spatial resolution*, or *scale*. It is not possible to get meaningful results from the combination of spatial data recorded to a scale of 1:250, as might be the case for an excavation site plan, with road alignments recorded to a scale of 1:250,000. In the former example 1mm represents 25cm, and in the latter example represents 250m. Spatial data recorded to *scales* of greater than around 1:10000 involve considerable generalisation of alignments to avoid features conflicting. This is especially true of *paper* maps drawn to such scales.
- Non-spatial information to be combined, or integrated, must use the same field definitions, encoding regimes, etc. Where different schemes are used it will be necessary to convert or translate the data to the required scheme.

2.1.4 Layers and Themes

The terms *layer* and *theme* are used almost interchangeably by many people - archaeologists and GIS practitioners included - yet are given very distinct meanings by some software suppliers and in some specific disciplines, for example in Computer Aided Design (CAD). For the purposes of this guide these terms will be used as follows. A **theme** is a collection of like objects, for example 'pottery', 'Iron Age sites', etc. A **layer** is a group of specific objects within a *theme* - for example, 'Stamford Ware' within the pottery theme or 'Hillforts' in the Iron Age site theme. In order to avoid confusion, it is important that the names given to such themes and layers are both descriptive and free from ambiguity.

Figure 2: Layers and Themes - Figure created by Peter Halls using data from the Cottam Project directed by Julian Richards. Image copyright © Archaeology Data Service. Two layers from the Cottam Project GIS (red layer with coin findspots and black layer with metal artefact findspots) are combined to represent the metal object theme. Two other themes are illustrated in this image, one for aerial photograph interpretations (line data) and another for geophysical survey data (raster image).

The purpose of the theme/layer approach is to provide a framework for collecting together objects of similar nature - in terms of either representation and/or descriptive type. Thus, different Iron Age site types might be gathered together because they are related in terms of both the representational type -- a line or point object -- and because of their nature or purpose -- delineation of the landscape location selected for settlements in the Iron Age. In the same way, a database of finds of pottery might be defined in locational terms as a collection of points, each of which might relate to an individual object, or closely related group of objects.